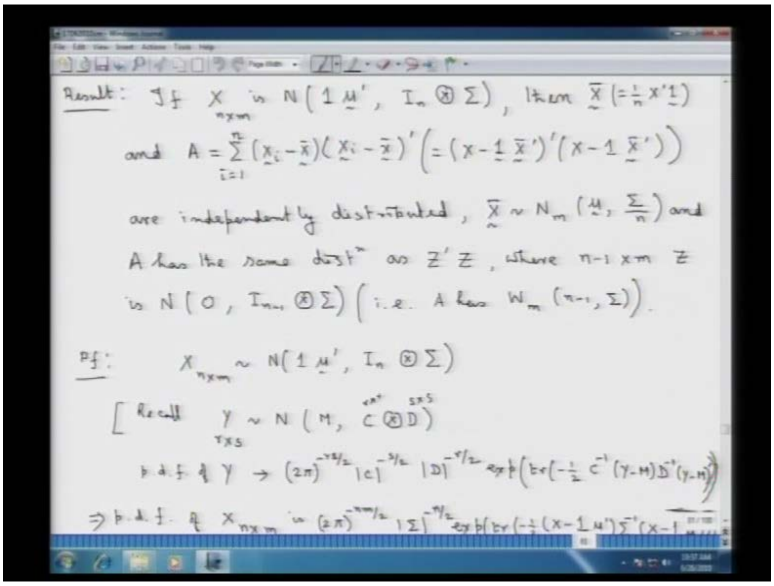
Applied Multivariate Analysis
Prof. Amit Mitra
Prof. Sharmishtha Mitra
Department of Mathematics and Statistics

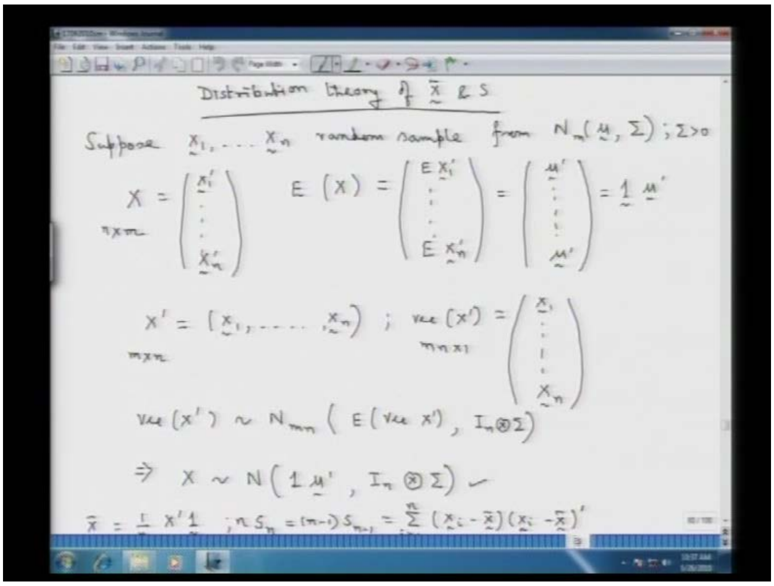
Indian Institute of Technology, Kanpur
Lecture No. # 11
Wishart Distribution and it's Properties – I

(Refer Slide Time: 00:21)



So, let us recall, what we were doing in the last lecture. We were proving an important result in the multivariate distribution theory, when we have a random sampling from a normal distribution, multivariate normal distribution.

(Refer Slide Time: 00:33)

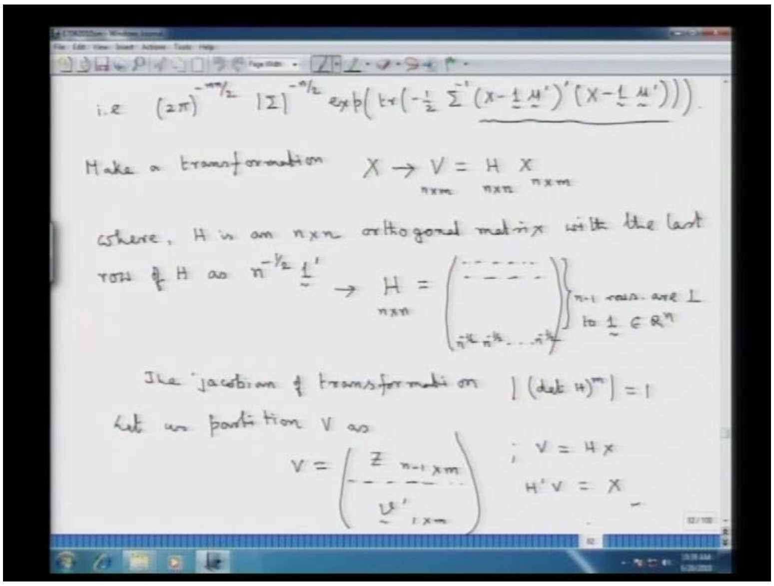


So, from the multivariate normal distribution sampling, what we had was X 1, X 2, X n a random sample from multivariate normal with a mean vector mu, and covariance matrix sigma. So, using this set of random vectors, which are random samples from this multivariate normal distribution. We had formed this data matrix X, which was n by m, and we had that the distribution of X to be a matrix normal distribution given by the following parameters with a mean matrix as 1 mu prime, and the covariance matrix of vec of X prime was I n Kronecker product sigma. So, under such a setup, we had a stated that this result, and started proving that in the last lecture, that we have X bar, the sample mean random vector to be given by 1 upon N X prime one, and the matrix A which is a constant multiplier of the sample variance, covariance matrix in either of the forms with a divisor n or a divisor n minus 1.

So, we had this quantity of interest, and the result basically, tells us that X bar, and A are independently distributed; X bar having a multivariate normal distribution with these parameters with mean vector as mu, and covariance matrix as sigma by n. And A has got the same distribution as that of Z prime Z, where the n minus 1 cross m Z random matrix is a matrix normal distribution; with mean matrix as a null matrix, and a covariance matrix of vec of Z prime to be I n minus 1 Kronecker product sigma.

That is in other words, A has got a Wishart distribution, m dimensional with parameters, degrees of freedom as n minus 1 and the associated variance, covariance matrix as sigma. So, we started with probability density function of this X random matrix, which we had written it is in this particular form.

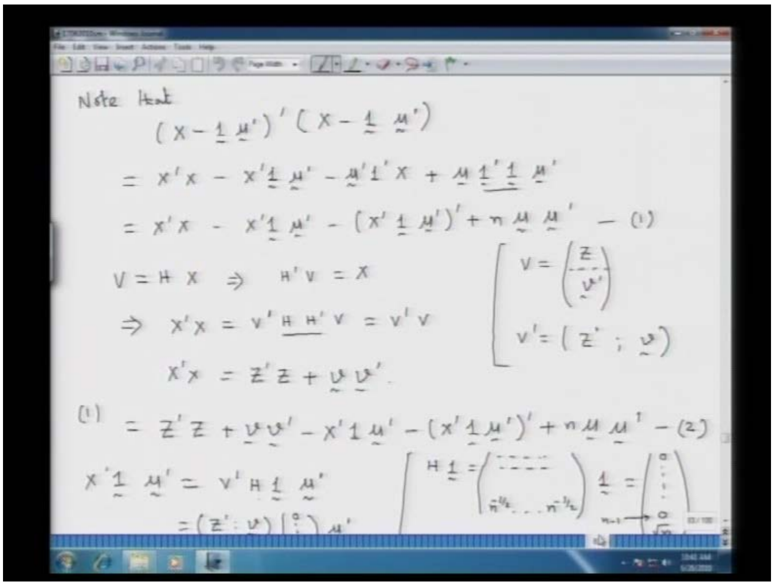
(Refer Slide Time: 02:27)



Then, from this form of the probability density function of the random matrix X, we had made a transformation from X to V; V was given by H times X, and H was an orthogonal matrix with the last row of H, given by n to the power minus half in all the positions. So, with this transformation we were trying to see, what is the joint p d f of this transformed random matrix? Which is V n by n; so, this H matrix plays a crucial role. The type of H matrix that, we had taken in the special form. That of course, had implied that the n minus 1 rows; the first n minus 1 rows, this is the first row, second row, and n minus 1th row. All these, n minus 1 rows of H are orthogonal to this 1 vector; 1 vector is a n dimensional column vector with 1 at all the positions. The jacobian of transformation was seemed to be equal to 1.

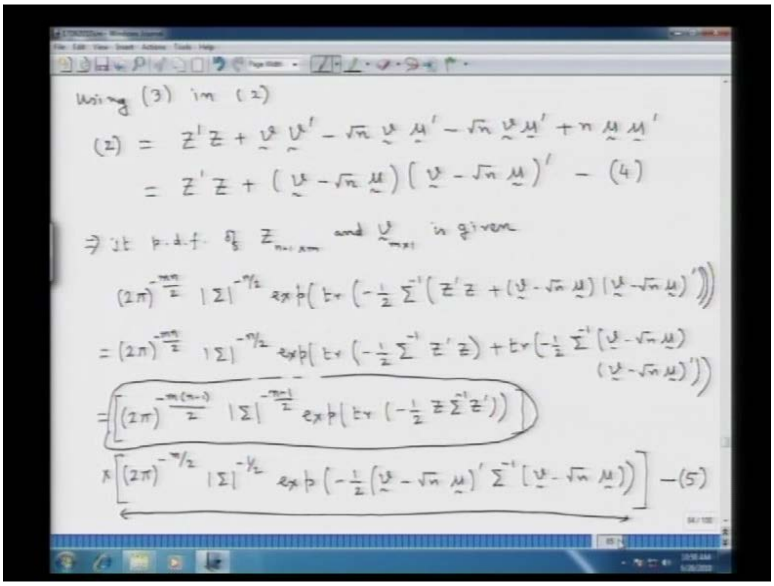
Now, we had partitioned with a purpose of course; this V, the new set of random variables forming this random matrix into this Z, and V prime Z was an n minus 1 cross m random matrix, and this V prime was a row vector m dimensional. So, from this we were trying to see, what is the joint p d f of this V, of this Z, and V transpose?

(Refer Slide Time: 03:53)



So, in order to do that, we had these calculations done; we had seen, that this exponent part, which was appearing here. The exponent part in the p d f, which is (( )) this trace of minus half X minus mu transpose sigma inverse X minus mu transpose whole transpose. So, this particular quantity or in other words, that was written in this form. That it is E to the power trace minus half sigma inverse X minus 1 mu transpose whole transpose into x minus 1 mu transpose. So, we had obtained, what is this quantity; which is in the exponent X minus 1 mu transpose transpose X minus 1 mu transpose in terms of the newly defined random variables.

(Refer Slide Time: 04:45)



So, that we had up to this particular point that this (( )), which was here two term, which is this term? Here is basically equal to two is, what was sitting in the exponent of that joint p d f. So, from there we had the form of two written in this particular form. So, this would imply now, that the joint p d f of this Z matrix, which is n minus 1 cross n, and V. Now, this V is n by 1 vector is given by now, the constant part remains as it is so, this is minus m n by 2 determinant of sigma to the power minus n by 2, and then, we had in the exponent E to the power trace minus half of that X minus 1 mu prime whole transpose into X minus 1 mu prime. Then, that term is basically equal to this term. So, we will have that along with that sigma inverse matrix, which we have not disturbed.

So, this is that sigma inverse matrix, this multiplied by Z transpose Z, plus V minus, root n times mu into V minus root n times mu transpose. So, this is corresponding to this, we will have two more brackets to close this particular expression. So, this is what we have as the joint p d f of this random matrix Z, and the random vector, which is V. Now let us, write how we can write this particular joint p d f of the transformed random variables in this setup. These constants at the moment are kept as it is, and we will have this the first term as trace of minus, half sigma inverse, Z transpose Z. So, that is the first term, and then we will have with trace here, trace of minus half sigma inverse V minus root n times mu. This multiplied by V minus root n times mu transpose. So, that this trace comes here, this trace is here we will require one more bracket. So, we have actually partitioned this particular joint p d f into two parts; one corresponding to this Z random matrix, and the other one, that is corresponding to the random vector; which is V. Let us, write these constants accordingly so, that we can have well defined distributions corresponding to these partitions.

So, we will write this, as m into n minus 1 by 2. Similarly, we will write this determinant of sigma to the power n minus 1 by 2. Then, we will have exponent trace of minus half Z sigma inverse Z transpose. There is no problem in writing it, because what we are doing is trace of a b equal to trace of b a. So, we can think this Z in the front. So, we will have this as a first set of terms, this multiplied by whatever constants are remaining. So, that this would thus be, we will have to adjust for this n minus 1 by 2, which we will have taken from n by 2.

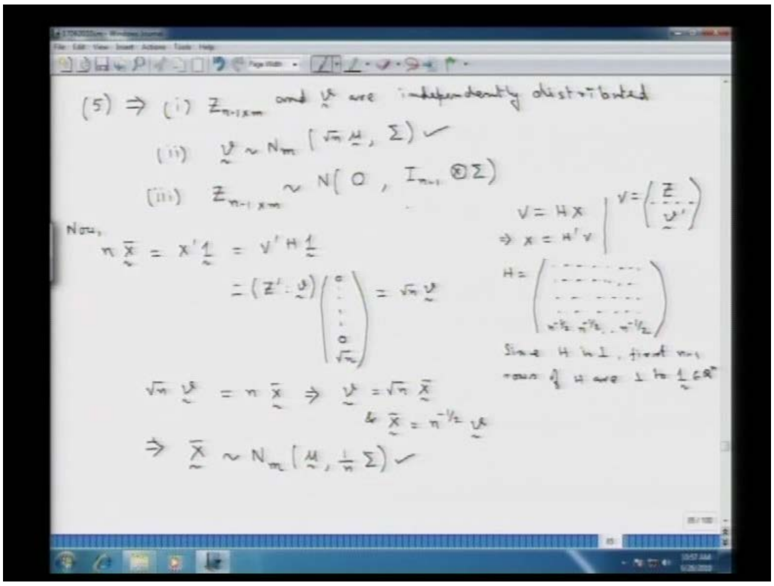
So, what we will have here is 2 pi to the power minus n by 2, because this is minus m n by 2 is already remaining, what we have added is m by 2. So, we make adjustment for that, and then we will have this as determinant of sigma to the power minus half, in order to adjust this particular term. Here, let me put a bracket here. So, this is the first expression that multiplied by the second expression, and what do you have in the second expression. The second expression for exponent is E to the power trace of this particular quantity.

So, we will have that written as E to the power minus half. Now, once again realize that, we are trace of these two quantities. So, we can write that trace of a b equal to trace of b a. So, we can take this term in front, and we will write this as V minus root, n mu transpose sigma inverse V minus root n times mu. So, that we have been able to partition this into two parts. Now, what is important to realize from this particular partition (( )) three important things. Now, what we have this is say a function of Z here. So, this is the density function or the joint density function for this random matrix Z, and this is what is a partition corresponding to this V vector.

So, we have the joint p d f of Z, and we written in terms of the product of the probability density functions of this random matrix Z. That is the first expression here, and the probability density function of the random vector, which is V; and hence we can say that, this random matrix Z and the random vector V are independently distributed. So, that is the first thing that, comes out from this writing the joint p d f of Z, and V. In terms of this particular product, the two other things; that emerge is that, if you look at this particular density; the second one, it is clearly the density of the multivariate normal distribution.

So, if we consider V, we clearly has got a multivariate normal distribution with what parameters, now the dimension of V is m. So, this v is going to have a multivariate normal m dimensional with a mean vector as this root n mu. So, that is what is the mean vector corresponding to this v, and what is the covariance matrix of this v vector. The covariance matrix of v is clearly, this matrix which is sigma. So, we have written it, in that particular form, which readily actually gives us idea about, what is what is the distribution of this associated V; and also note that, if we have this as a first part; which is corresponding to the probability density function of the random matrix Z. Then this reminds us of the probability density function of a matrix normal distribution. So, this basically is the density functions of this part. The first part here; so, this first part is, what is corresponding to the probability density function of a matrix normal distribution, and this second part here is corresponding to the probability density function of a multivariate normal distribution. So, we can actually put these things together, and write it as conclusion.

(Refer Slide Time: 12:21)



So, this would imply this means basically this particular expression. So, suppose we have this numbered as number 5. So, this expression 5 would imply, because that is a joint p d f of this random matrix Z, and the random vector V. So, that we have obtained it as the product of the two respected probability density functions. So, this would imply number one, that this Z n minus 1 cross m random matrix and V are independently distributed. So, that is an important realization, which would eventually lead us to proving that X bar and n minus 1 S or A are independently distributed; furthermore, this V is seen to have a m dimensional normal distribution with a mean vector, as root n times mu, and a covariance matrix as sigma, and what is the third conclusion, that we can draw from that expression. We can also say that, this random matrix Z n minus 1 cross n has got a matrix normal distribution with what would be the mean matrix. It is easy to see that, this is not centered around any mean matrix, and hence the mean matrix is a null matrix.

So, this has got a matrix normal distribution with null matrix as it is mean matrix, and what would be the covariance matrix? The covariance matrix would correspond from this particular expression; as you can see that, if we had a matrix normal distribution y with a mean matrix as n, and a covariance matrix of vec of Z prime as C Kronecker product d. Then here, what we will be having is determinant of c raise to the power of the order of d that multiplied by the determinant of d raise to the power of the C matrix. So, here what we have is determinant of sigma only present here.

So, that C Kronecker product d we will have C as an identity matrix, and then we will have d as sigma matrix, and then if you look at the order of the I matrix; what we will be having is going to be n minus 1. So, we have we can complete this particular statement. That this is a matrix normal distribution with I n minus 1 Kronecker product sigma, as it is variance covariance matrix of this vec of Z prime quantity. So, these are important things to note here. Now, we will use this three conclusions, that we have obtained from the joint p d f of Z, and V in order to prove our desired results.

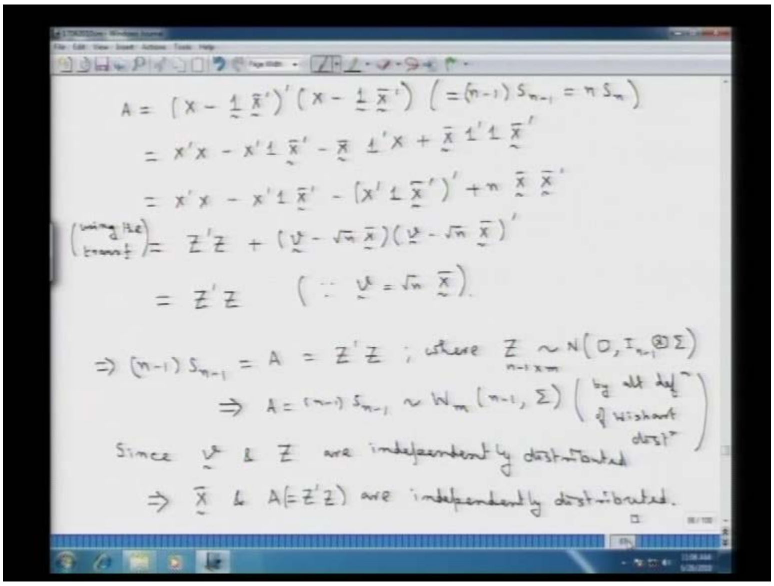
Now, recall that n times X bar is X transpose 1 vector, and what is that equal to what was the transformation that we had made we had made a transformation V; which was equal to H times X. So, this would imply that our X matrix; the original X matrix is H transpose V. So, that we can write in place of X transpose A V transpose H 1 vector. Now, we have already noted that from the structure of H, what we have? Let us, recall that also we had said that, H is an orthogonal matrix such that the first n minus 1 rows whatever they are have to be orthogonal to the nth row, which has got this special form that all the entries. In that, nth row are n to the power minus half.

So, since this H is orthogonal first n minus 1 rows of H the first n minus 1 rows of H are orthogonal to this 1 vector belonging to R to the power n, and hence when we multiply this H matrix with a vector of (( )) n dimensional. The first n minus 1 entries now note that this V. We had taken to be of the partition that we had taken this V as a partition, which was Z and V prime. So, that this v transpose would be Z transpose V augmented. So, we will have this as Z transpose augmented with this V vector, and then this H times 1 is going to give us this vector, which is zero on the first n minus 1 entries, and the last entry is going to be this column, this row rather multiplied by A 1 column.

So, we will have n multiplied by n to the power minus half, and that is just equal to this term, and thus this n times X bar which is in terms of the original random variables from the random sampling from a multivariate normal distribution; that in terms of the transformed random matrix V, and it is partitions Z, and this V prime is nothing, but root n times V. So, that this would now imply that, these root n times V is equal to n times X bar. That is, this V is root n to the m to the power minus half. So, that would be root n times this X bar vector. Now, we know what is the distribution of V? We have already obtained that, and hence from there we can derive, what is the distribution of this X bar. So, this X bar would imply from this observation; using this, that V has got this multivariate normal m root n mu sigma matrix. So, we will have root n times V, which is going to be n times X bar will have root n times.

This as n times that, and X bar which is going to be this divided by n. So, that would be n 1 upon root n times V. We will also have from here, and this X bar from here. So, we are dividing the two sides by n. So, we will have this as n to the power minus half times this V vector. So, this would imply that, X bar will follow, because it is just a constant multiplier of this V vector. So, multiplying that by n to the power minus half the mean vector; the dimension remains the same. So, the mean vector would just be equal to mu, and then we will have one upon n times sigma as the variance covariance matrix. So, this proves one part of the result only. Now, let us look at how to use this part, and this part in order to prove the remaining portions of this result.

(Refer Slide Time: 20:11)



Now, note that the A matrix; that we had defined was given by X minus 1 X bar transpose whole transpose into X minus 1 X bar transpose. So, we have this as our A, we can write this term by term as X transpose X this minus X transpose 1 X bar transpose; then, we will have from here a minus just the transpose of this nothing else. So, X bar transpose sorry this is just X bar, because we have got this as X bar.

So, we will have X bar 1 transpose this X, and then with a plus sign, we will have this X bar 1 transpose 1 X bar. So, this is equal to X transpose X minus X transpose 1 X bar transpose this minus X transpose 1 X bar transpose whole transpose. So, this term is just the transpose of this quantity, this plus n times, this is a transpose; here n times X bar X bar transpose. Now, we can simplify this particular term exactly in the same way, as we had simplified a similar expression with X bar being replaced by mu. Let us, see what we had obtained there.

Now, see this expression, and the expressions that we are considering now are similar in nature. Basically, it is different here in the present expression; what we have is this mu is replaced by X bar, and there is no other difference. So, we had that term finally, reducing to this term. So, that we had that being reduced to this Z transpose Z, plus V minus root n times mu into V minus root n times mu transpose. Now in the present expression, if we just replace mu by X bar; we are going to have the form of the expression which, we are now looking at this term is nothing, but we will have that as Z; this is Z, Z transpose using the transformation. Here, using the transformation, and proceeding exactly as in the previous expression, where we had in place of X bar just that mu, and this plus V minus root n, we had previously root n times mu. Now, mu is not present here, we what we have is X bar. So, that would be V minus root n times X bar; this let me see what it was exactly so, that was Z transpose Z in to this transpose. So, we will have this as a Z transpose Z, and then we will have this as V root n X bar transpose.

So, this is, what is A? The constant multiplier of the variance covariance matrix so, this is just to recall, this is n minus 1 times s n minus 1 or that is n times S n the sample variance covariance matrix in two different forms. So, what can we say now, about this particular expression note that, what we have obtained is that V equal to root n times X bar. So, since V is equal to root n times X bar v is equal to root n times x bar this expression vanishes. So, this is just Z prime Z this is, because V by way of construction is root n times this X bar.

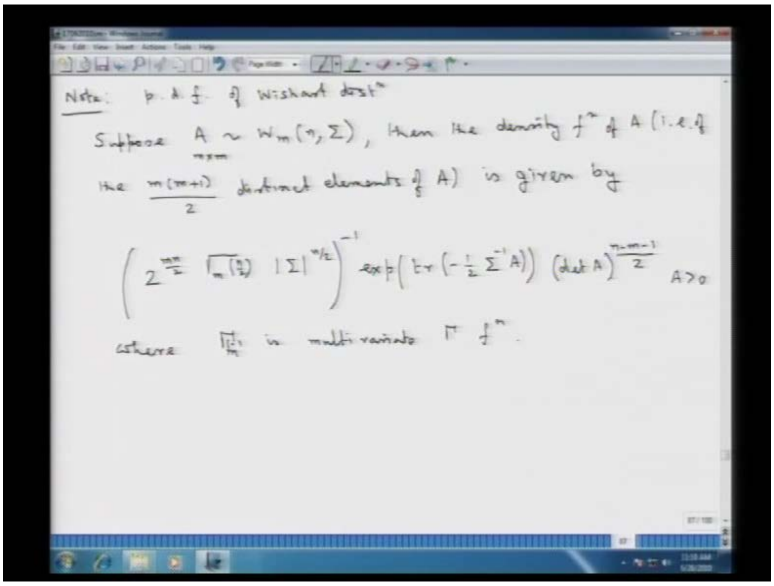
So, what we have obtained this A? Which is n minus 1 times S n minus 1; that is, A is equal to Z transpose Z, where this Z n minus 1 cross, n is A matrix normal distribution with a mean matrix as a null matrix, and a variance, covariance matrix of vec of Z prime is I n minus 1 Kronecker product sigma. So, what does that imply that, simply implies that this A has got the distribution, which is same as that of Z prime Z, where Z has got a matrix normal distribution with a null matrix as a mean matrix, and I n I n minus 1 Kronecker product sigma as the variance, covariance matrix of vec of Z prime.

Now, from the the alternate second alternate definition of the Wishart distribution, that we had given thus this A follows a Wishart distribution. So, this would imply that A, which is n minus 1 times S n minus 1 will follow a Wishart distribution n dimensional on n minus 1 degrees of freedom, and an associated variance, covariance matrix of sigma. This follows by alternate definition of the Wishart distribution. So, what have we proved, and what have what we have not yet proved. Now, we have already proved that, X bar has got that multivariate normal distribution; we have also proved that A n minus 1 S n minus 1 or equal to n times s n has got a Wishart distribution, Wishart m n minus 1 and sigma.

Now, (( )) that since this is the last part. So, since we have V, and Z are independently distributed. These two are independently distributed; this would imply now X bar is coming from V, and A is coming from Z, and hence X bar, and A are going to be independent. So, this would imply further that this X bar and A; now, X bar is in terms of this V vector and A is in terms of this Z matrix. So, that X bar, and A are independently distributed. So, that concludes this particular proof of this very important result in multivariate distribution theory. So, that lets once again look back at the result stated.

So, we had to prove that X bar, and A, this the constant multiplier of the variance, covariance sample variance, covariance matrix are independently distributed. So, we have proved this; we have proved that X bar has got multivariate normal distribution mu sigma by n. We have also proved that A has got the same distribution as that of Z prime Z, where Z is a random matrix; which is having a matrix normal distribution zero I n minus 1 Kronecker product sigma. That is, in other words this has got a Wishart distribution on m n minus 1 degrees of freedom.

(Refer Slide Time: 28:22)

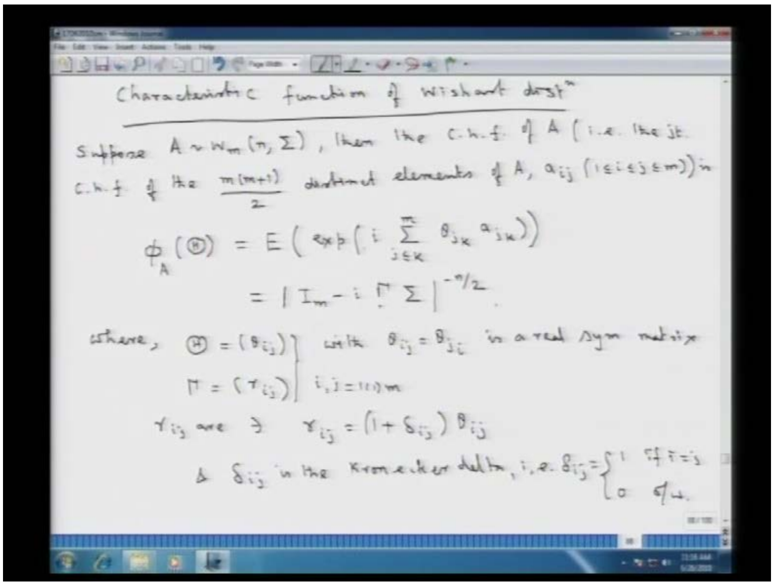


Now, let us put as a note we are not going to derive this; the density function of a Wishart distribution, how it looks actually, because see we have not used this particular p d f, but for the sake of completion. We will just write this the p d f of a Wishart distribution. So, p d f of this Wishart distribution now, suppose we have A random matrix, A suppose that is m by n suppose this follows a Wishart distribution m n sigma. Now, note that when we say that we have a Wishart distribution m by n from the definition of the Wishart distribution; either 1 says that, Z will follow a Wishart distribution, when Z is or rather A is said to have a Wishart distribution m n sigma. If A can be written as summation i equal to 1 to n Y i Y i transpose where Y is an independent multivariate normal distribution with a mean vector as null vector, and a covariance matrix as sigma matrix.

So, that is the first definition that we had given or alternatively we can say that we will say that A follows a Wishart distribution m n sigma. If A is given by Z transpose Z where Z has got a matrix normal distribution. So, what are the V? The definition that we consider, it is easy to see that this matrix K is a symmetric matrix, because either you write A as the Wishart matrix as summation Y i Y i prime, which is symmetric or you writes A as Z transpose Z; where Z has got matrix normal distribution; once again that also is symmetric. So, when we talk about p d f of the Wishart distribution. It is basically, we are looking at the joint p d f of the distinct elements that are present in this A matrix.

Now, how many distinct elements are present in this A matrix; this A matrix is a symmetric matrix. So, the number of distant elements are m into n plus 1 by 2. So, that then we are talking about the probability density function of a Wishart distribution; we are essentially talking about the joint probability density function of the m into m plus 1 by 2 distinct elements of this wishart matrix. So, let me write this then the density function of A, that is of the m into m plus 1 by 2 distinct elements of A is given by the following expression; I will just write it. So, this is 2 to the power m n by two; this is a multivariate gamma function multivariate gamma of order m of n by two, then determinant of sigma to the power n by 2 whole raise to the power minus half, then we have exponent of trace of minus half sigma inverse A (( )) into determinant of A whole raise to the power n minus m minus 1 this divided by 2; now, for A to be positive definite; this is what is the probability density function or the joint probability density function of the m into m plus 1 by 2 distinct elements of this Wishart distribution; where this function is a multivariate gamma function; where this gamma this dot is multivariate gamma function. So, although we have not used this particular density in order to derive any results as such for the sake of completion; this is what the density function of a Wishart distribution.

(Refer Slide Time: 32:56)



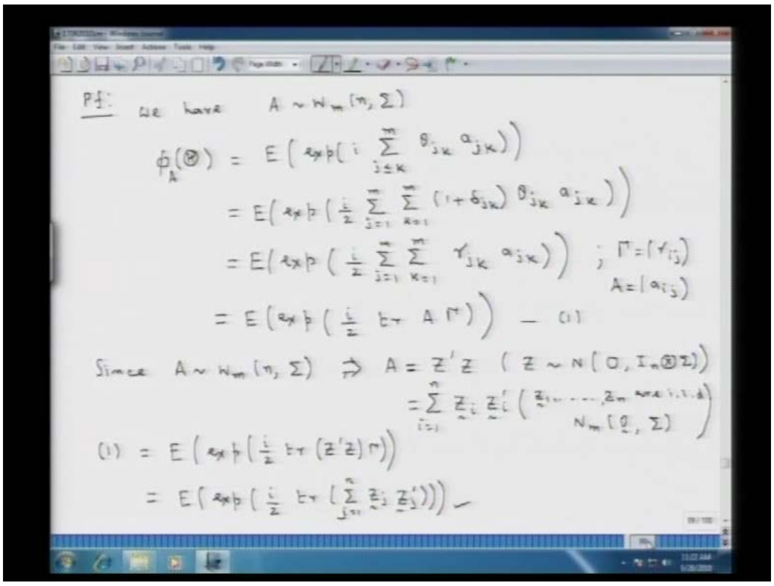
Now, the next thing that we are going to look at is the characteristic function characteristic function of Wishart distribution. Let us look at, how this characteristic function actually looks like; let me, state the result first; suppose, we have A following a wishart m n sigma. Then the characteristic function of A, now once again this is a joint characteristic function of the district elements; that is the joint characteristic function of the m into m plus 1 by 2 distinct elements of this random matrix A; let me, write then as some notations of A say a i j.

Now, for the distinct elements, what we will be looking at is one direction only. So, these are basically the random elements, what we have? So, the characteristic function of this is given by the following; let us, denote that by phi A of at the points, this script theta matrix which is given by expectation of E to the power i times summation, double summation j less than or equal to K equal to 1 to up to n. Then, we have this as theta j k times a j k; that is basically is the joint characteristic function, because we are looking at all the distinct elements. So, that this would just be given by this j less than or equal to K of these random quantities. In this double summation, we are only looking at the distinct elements in one direction.

So, we have that to be equal to determinant of an identity matrix of order I m minus i times gamma times, this sigma matrix, and whole raise to the power minus n by 2. So, this is what is the going to be the characteristic function of a Wishart distribution, where we have these notations; that we have already introduced, where script theta is a matrix; which is having elements as theta i j. It is a symmetric matrix with theta i j equal to theta j i is a real symmetric matrix. So, this we have a real symmetric matrix, and this matrix gamma. That, we have defined here is having elements say gamma i j, i j for both these quantities i j equal to 1 to up to m.

Now, this gamma i j are such, that this gamma i j is equal to 1 plus delta i j into theta i j, and this delta i j is a Kronecker delta is the Kronecker delta; that is this delta i j is equal to one, if i is equal to j and is equal to 0, if it is otherwise so, this what is the statement of the characteristic function of Wishart distribution. If we have a Wishart distribution, wishart m n sigma; then the characteristic function of A which is also actually the characteristic function of the or the joint characteristic function of the m into m plus by 2 distinct elements of this matrix A; which are a i j, which is given by phi A script theta is going to be of this particular form; we now, look at proving this particular result. That is deriving that the characteristic function of the Wishart, really is given by this.

(Refer Slide Time: 37:24)

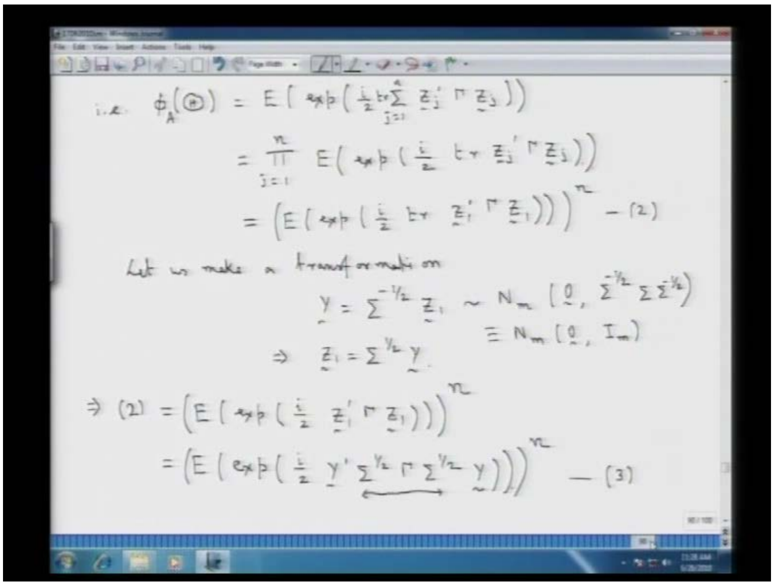


So, let us look at proving this important result, which gives us the characteristic function of a Wishart distribution. So, we have this A to follow a Wishart distribution, wishart m n sigma, and for that, what we have is this phi A script theta; which is the characteristic function is expectation of E to the power i times, double summation j less than or equal to k, equal to 1 to up to n which is theta j k times a j k. So, let us write this in the following form; that it is E to the power, we remove this particular restriction; we write it as half of this. So, we will have j equal to 1 to m k equal to 1 to m, and then we will have that in order to adjust. We will have a delta j k times, theta j k times a j k. So, if we remove this restriction that j is less than or equal to K in order to adjust that, what we have introduced is this Kronecker delta j k, and we have written it in this particular form.

Now, this form will lead us to now note that this particular term here, 1 plus delta j k into theta j k is nothing gamma j k. So, we will have E to the power i by 2 summation j equal to 1 to n, summation K equal to 1 to n, and then this is gamma j k, the notation that we are introduced that gamma j k is 1 plus delta j k times theta j k. So, we will have that to be given by this; now, what is this expression by the way, if we look at this double summation here, j equal to 1 to n k equal to 1 to n theta gamma j k times a j k. So, gamma is that matrix, which is holding this gamma i j terms, and A is the matrix, which is holding this a i j terms. So, if we look at this double summation here. It is nothing, but trace of the product matrix gamma with A.

So, we can write this as E to the power i by 2. Then, we have trace of A times gamma matrix. I can write it as A times gamma or you can write it as gamma A, because we have a trace here. So, does not matter; now let me, write this as equation number 1. In this proof, now since we have A to follow a wishart m n sigma; we can write A as Z transpose Z either where this Z has got A matrix normal distribution, this has got wishart m n sigma. So, we will have this as A null matrix, and i n Kronecker product sigma, as it is variance, covariance matrix or we can write this. So, these are the two alternate definitions, alternate equivalent definitions of the Wishart distribution that we had given one can write this as summation Z i Z i prime i equal to 1 to n; where Z 1, Z 2, Z n, are i i d independently, and identically distributed multivariate normal random vectors with mean vector as null vector, and a covariance matrix as sigma. So, these two are known facts; now, this first will lead 1 to the following, we can write this as expectation of E to the power i by 2 trace of in place of A, we can write this as Z transpose Z times. This gamma matrix or we can write that, equivalently in terms of this Z i vectors as expectation of E to the power i by 2 trace of summation i. Let me have this as j, because we have i is the imaginary part; which imaginary number, which is square root of minus 1. So, we have j equal to 1 to up to n Z j Z j prime of this particular term; it would be advantages, actually to use this particular form in order to prove this result. So, let me use this particular form.

(Refer Slide Time: 42:34)



That is we have phi A script theta, that is given by expectation of E to the power the expression. There was i by 2 one can write this as summation Z j prime gamma Z j, j equal to 1 to up to n, how have we obtained this particular expression from the previous expression, if we look at this; so, we can take the trace. There is a gamma matrix somewhere, which is slipped out, this is that gamma matrix.

So, we will have here, once we have this is trace of this Z j quantity, and that multiplied by this gamma, and then we have the brackets closing; this is first bracket closing here, and this is the exponent bracket closing here. So, we will have trace of this particular term. Here, now the gamma can be multiplied out here term by term. So, we will have trace of summation Z j Z j transpose times gamma, and then trace of a b equal to trace of b a. So, what will be having here is trace of this Z j Z j transpose gamma, will be equal to trace of Z j transpose gamma Z j, and that is what is written out here.

So, we will have that expression as i by 2 remains, outside summation i equal to 1 to n Z j transpose gamma Z j. Now, here this is Z j are i i d multivariate normal random variables; each with a normal distribution with a mean, a vector zero, and a covariance matrix as a sigma matrix. So, since this Z j are independent; one can write this in terms of this product well without that this j equal to 1 n. So, we are looking at these expectations term by term; so, expectation of E to the power i by 2. There is a trace sitting here; i by 2 trace of 1 of these quantities; that is this Z j transpose gamma times this Z j.

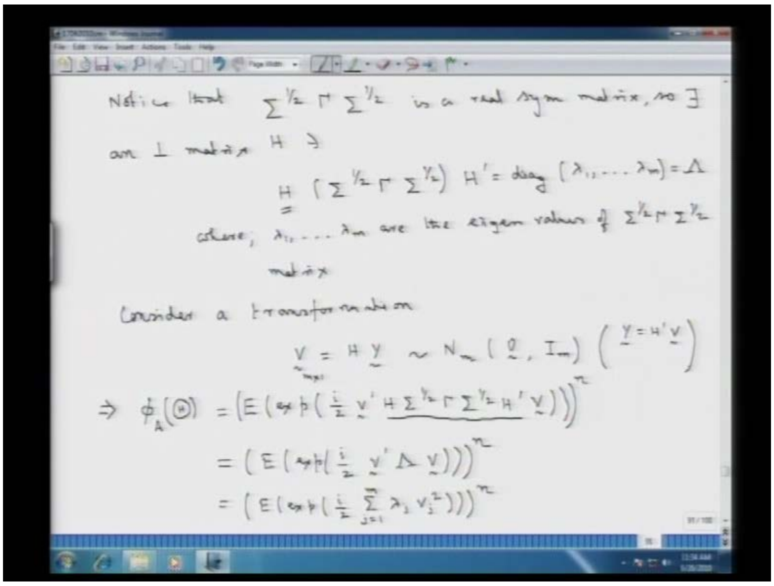
Now, note that this Z j are all i i d random vectors, and hence whatever the expectation for one Z j. That would be the same for the other Z j terms, and hence we can write it say expectation of E to the power i by 2 trace of Z j transpose gamma Z j. So, say let me just write this as 1 of these Z is say, I write this as Z 1. So, whatever is the expectation of this term; here, expectation of this exponent for Z 1 is going to be the same for any of the j; which is in this product here, why is that. So, because Z are i i d random variables, and hence this expression would just be raised to the power n nothing else.

Now, we have (( )) this particular expressions; now, let us make a transformation let us make a transformation say Y, which is equal to sigma to the power minus half times this Z 1 vector. Now, remember Z 1 has got a multivariate normal mean vector zero, covariance matrix sigma. So, this Y will follow, once again a multivariate normal distribution with the same dimensional T as that of the Z 1 vector. So, it would be this with a mean vector as a null matrix, null vector, and the covariance matrix as sigma to the power minus half covariance matrix of Z 1; which is sigma times sigma to the power minus half transpose. So, it is a symmetric matrix. So, this is equal to this.

That is, what we have is this a multivariate normal distribution m, and this is just an identity matrix, but we can replace this Z is by the corresponding y i terms. Now, if Y is given by sigma to the power minus half Z 1; this would imply that Z 1 random vector is this being pre multiplied by sigma to the power plus half. So, we will have this sigma to the power plus half into this Y vector. So, let me have this as number 1. So, this will imply that this number 2. I am sorry this is number 2. So, this number 2 is expectation of this is raised to the power n. So, expectation of E to the power i by 2 then trace of Z transpose Z 1 transpose.

So, I can just remove this trace as well, because this is a scalar quantity out here. So, that we can this is this is Z 1 transpose gamma Z 1; this is a scalar quantity. So, we can either write that as trace or drop that trace, because this is now going to be a scalar quantity. So, what we have this as i by 2 Z 1 transpose gamma times. This Z 1, this raised to the power n; now, this in terms of the y random vector would be expectation of E to the power i by 2. So, Z 1 transpose would be Y transpose sigma, half gamma, sigma half times this Y vector. So, this bracket closes here, and this bracket closes here, this is for this this is for this and we will have this raised to the power n. So, the characteristic function, thus is of this particular form; now, let us try to see what is the what is that special about this particular matrix sigma, half gamma, sigma half. Let me, give this equation number 3, and move on to realizing what is this.

(Refer Slide Time: 49:24)



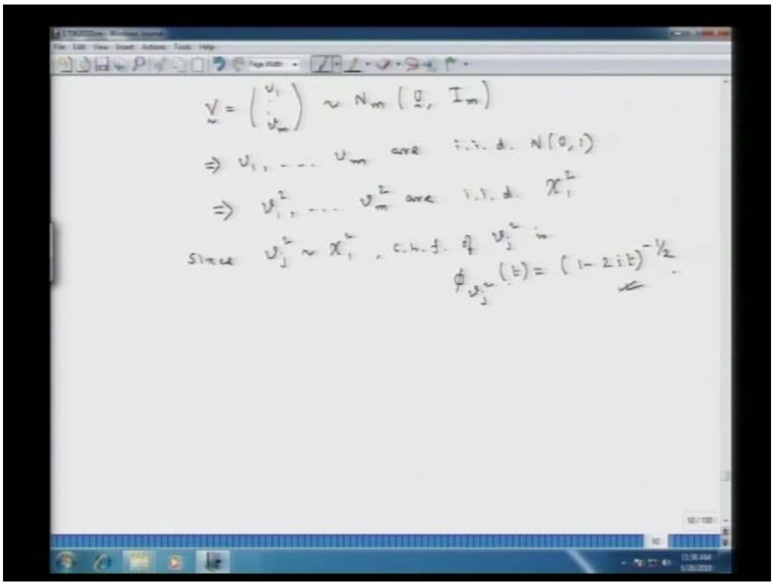
We notice that this sigma, half gamma, sigma half is a real symmetric matrix. So, there exist an orthogonal matrix H; say such that, we will be having this H sigma, half gamma, sigma half H transpose to be given by a diagonal matrix lambda 1, lambda 2, lambda m. Let us, denote that by a capital lambda, where this is? Basically, the spectral decomposition of this real symmetric matrix; where lambda one, lambda two, lambda m are the eigen values of this sigma, half gamma, sigma half matrix, and what is H? H is that orthogonal matrix; which is having the orthonormalized eigenvectors corresponding to these eigen values of sigma, half gamma, and sigma half matrix. So, we will have this spectral decomposition corresponding to that matrix.

Now, using this matrix, orthogonal matrix H make a transformation; consider a transformation from Y to say a random vector V; which is given by H times Y, now H is this orthogonal matrix. We know, what is the distribution of Y? what is that distribution of y The distribution of Y is multivariate normal with a mean vector as null vector, and a covariance matrix as I m. So, this will lead us to, what is the distribution of this V; this V is going to be now, what is the order of this H matrix; H matrix naturally is m by n. So, this is an m by n. So, we will have this multivariate normal n with a mean vector as a null vector, and the covariance matrix as H, covariance matrix of Y; which is an identity matrix times H prime.

Now, H is an orthogonal matrix. So, we will have that, once again to be given by this I m is an identity matrix of order m. So, this would imply this phi A script theta is equal to what we will take what we will take forward is this particular form, and then in place of Y; what will be writing is V. So, we are making that transformation. So, that it would be a bracket is outside expectation of E to the power i by 2. Then, we have this now from here; this V is H times Y. So, our Y is going to be H transpose V. So, from here what we need to look at this, what is Y transpose. So, that it would be V transpose H, then sigma, half gamma, sigma half, this remains as it is; that is multiplied by Y vector and Y vector is nothing, but H transpose this V vector. So, this is what is there in the exponent; this term closes here, and this term closes here, this is raised to the power n. So, what we have here is that note that this particular term here; which is H sigma, half gamma, and sigma half; H transpose is nothing, but this gamma matrix. So, we will have this as expectation of E to the power i by 2 V transpose. Then, this capital lambda matrix there, because that is what is given here? That is multiplied by this V vector, and then that is raised to the power n.

So, what would that be equal to this would be equal to i by 2? Now, V is that particular n dimensional vector. Now, this capital lambda is this diagonal matrix, which is holding the eigen values lambda 1, lambda 2, lambda m. So, that this is just going to be equal to summation. This lambda j, v j square terms what are this v j? This j is equal to 1 to up to n; that is the order of this V vector this V vector is m by 1, and hence the elements are v 1, v 2, v m, and these basically are the square of those entries out here. So, that we will have this to be equal to this; now, note that what is the what is that special about this v i entries.

(Refer Slide Time: 54:49)



Now, this V vector which is our v 1, v 2, v m; this follows a multivariate normal m dimension with a mean vector as null vector, and a covariance matrix as I m. So, this would imply that, this v 1, v 2, v m are i i d normal 0, 1. So, that these v 1, v 2, v m are standard normal random variables. So, this would imply that, v 1 square, v 2 square, v n square, they are the squares of the standard normal variates; since v 1, v 2, v m are i i d, and so, will be v 1 square, v 2 square, v m square; these are i i d what random variables these are chi square on one degrees of freedom variables.

So, what we have obtained? What we have reduced; this cumbersome derivation of this deriving the characteristic function of the Wishart distribution; this we have derived in terms of the characteristic function of just a simple chi square random variate on one degrees of freedom. So, that we can say that this; since, this v j square has got a chi square on one degrees of freedom. The characteristic function of this v j square is given by phi v j square t that is equal to 1 minus 2 i t that raised to the power minus half.

So, we will use this characteristic function; now, each of these v 1 square, v 2 square, v m square, we are having the same characteristic function, because they are our d random variates, and hence the character i i d chi square, center chi square on one degrees of freedom random variates, and hence the characteristic function is going to be the same. So, we will use this characteristic function of the central chi square distribution, and this plug in the values of those here in order to get to the final form of the Wishart distribution that, we will see in the next lecture. Thank you.